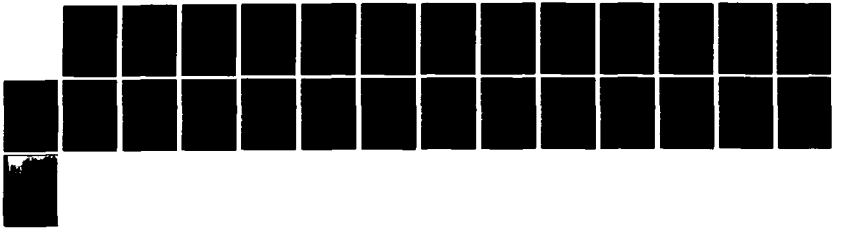
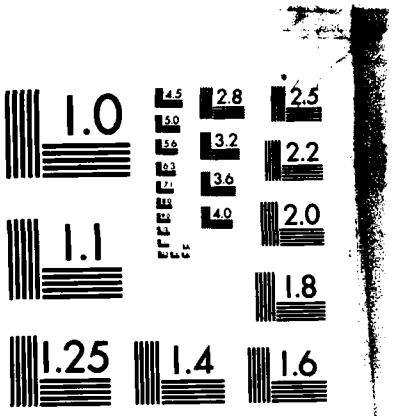
AD-A133 169 TWO-STAGE GROUP SCREENING SEARCH ROUTINE: PROGRAM DESCRIPTION AND USER'S GUIDE(U) DESMATICS INC STATE COLLEGE PA C A MAURO ET AL. SEP 83 TR-113-10 1/1

UNCLASSIFIED N00014-79-C-0650 F/G 12/1 NL





TWO-STAGE GROUP SCREENING SEARCH ROUTINE: PROGRAM DESCRIPTION AND USER'S GUIDE

by

Carl A. Mauro
and
Kevin C. Burns

— STATISTICS —

— OPERATIONS RESEARCH —

— MATHEMATICS —

DESMATICS, INC.

P.O. Box 618
State College, PA 16801

DESMATICS, INC.

P. O. Box 618
State College, PA. 16801
Phone: (814) 238-9621

Applied Research in Statistics - Mathematics - Operations Research

TWO-STAGE GROUP SCREENING SEARCH ROUTINE: PROGRAM DESCRIPTION AND USER'S GUIDE

by

Carl A. Mauro
and
Kevin C. Burns

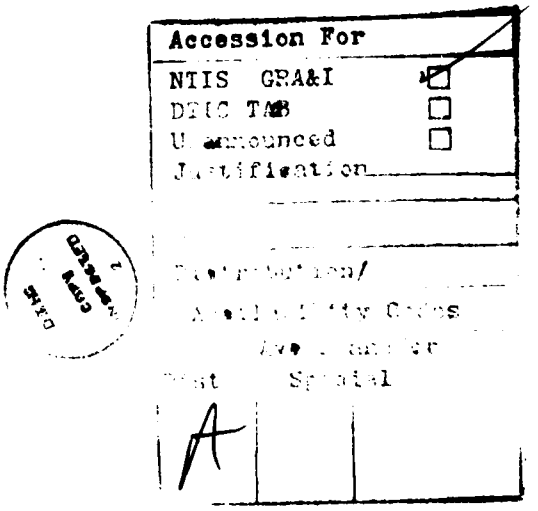
TECHNICAL REPORT NO. 113-10

September 1983

This study was supported by the Office of Naval Research
under Contract No. N00014-79-C-0650, Task No. NR-042-467

TABLE OF CONTENTS

		Page
1.	INTRODUCTION	1
2.	MODEL ASSUMPTIONS	2
3.	PROGRAM DESCRIPTION	4
4.	USER'S GUIDE	7
	4.1 IMSL Subprograms	7
	4.2 ERRSET Subroutine	7
	4.3 Input Considerations	8
	4.4 Program Modules	9
	4.5 An Example	9
5.	PROGRAM LISTING	13
6.	APPENDIX	18



1. INTRODUCTION

Many experimental situations require consideration of a large number of factors. In such situations there is often a need, because of resource limitations, for an efficient method of factor screening. One such method is two-stage group screening. In this method the individual factors (each at two levels) are partitioned into groups, forming group factors. By assigning the same level to all component factors within each group, the group factors are tested as if they were single factors. All factors in groups found to have a significant effect are then tested individually in a second-stage experiment.

In Desmatics Technical Report No. 113-9 (August 1983), "On the Performance of Two-Stage Group Screening Experiments," a methodology is developed with which to evaluate the performance of the two-stage group screening procedure. Also in that report, a computer-aided search routine is described. This routine may be used to help select a satisfactory group screening strategy. In the present report, we review the basic screening model, give a brief description of the search routine, and provide a listing of the program. The search program, which is written in standard FORTRAN, is easy to use, has intuitive appeal, and provides practical insight into the use and selection of a two-stage group screening strategy.

2. MODEL ASSUMPTIONS

We assume the following model:

$$y_i = \beta_0 + \sum_{j=1}^{K} \beta_j x_{ij} + \varepsilon_i$$

where K denotes the number of factors to be screened and $x_{ij} = \pm 1$ is the level of the j^{th} factor in the i^{th} run. We additionally assume:

1. $k \geq 1$ (k unknown) of the K factors are active (i.e., have a nonzero effect) and the remaining (K-k) are inactive,
2. all k active factors have the same <u>absolute</u> effect, $\Delta > 0$, and
3. the error terms $\{\varepsilon_i\}$ are i.i.d. $N(0, \sigma^2)$ random variables with σ^2 unknown.

In view of assumptions (1) and (2) we let $\underline{\beta}(i)$ for $i = 0, 1, 2, \ldots, k$ denote the case in which i active effects are equal to $-\Delta$, (k-i) active effects are equal to $+\Delta$, and the (K-k) inactive effects are equal to zero. The $\underline{\beta}(0)$ and $\underline{\beta}(k)$ cases correspond to the situation in which all active effects are in the same direction. In these cases there can be no cancellation of effects within a group. For $\underline{\beta}(i)$ cases with $i \neq 0$ and $i \neq k$ cancellation can occur. The search routine which we have developed considers the $\underline{\beta}(0)$ and $\underline{\beta}([k/2])$ cases, the latter being the case in which the chance of cancellation is greatest, among all $\underline{\beta}(i)$ cases.

In the two-stage group screening strategy considered here, we assume that the K factors are partitioned randomly into G groups of size g; if K is not a multiple of g, we assume that the group sizes are taken as "evenly" as possible. To analyze the results of the first

and second stages, we use the multifactorial designs of Plackett and Burman (PB). We denote the levels of the significance tests performed at the end of the first and second stages by α_1 and α_2, respectively. Following the notation used in Technical Report No. 113-9, we denote such a strategy by GS(g, α_1, α_2).

3. PROGRAM DESCRIPTION

We can define three separate measures of performance. These are:

Power. We denote by A the number of active factors that are detected correctly, and we define $E_A = 100E(A)/k$ as a percentage measure of the power of a GS strategy to detect active factors.

Type I Error. We denote by U the number of inactive factors that are declared active, and we define $E_U = 100E(U)/(K-k)$ as a percentage measure of Type I error.

Relative Testing Cost. We denote by R the total number of runs required by a GS strategy, and we define $E_R = 100E(R)/B(K+1)$ as a percentage measure of expected relative testing cost where $B(K+1)$ denotes the number of runs required by a PB design for $(K+1)$ factors.[1]

In general, an increase in α_1 or α_2 will increase both E_A and E_U. An increase in α_1 will also increase E_R, although E_R does not depend on α_2. It is clear that trade-offs between E_A, E_U, and E_R must be made by the experimenter in order to select a "good" $GS(g, \alpha_1, \alpha_2)$ strategy. The appropriate trade-offs would most likely be determined by economic and operational considerations.

A general practical approach to this problem is to maximize

[1] PB designs exist for numbers of runs that are multiples of four only. Accordingly, one can define $B(x) = x+4-x(\text{mod } 4)$. We use $x = K+1$ instead of $x = K$ in our definition of E_R to guarantee at least one degree of freedom to estimate error.

power (E_A) subject to fixed constraints on relative testing cost (E_R) and Type I Error (E_U). This approach is the basis of the search routine. That is, in order to apply the search program, the user must first input values of K, k, and Δ/σ, the signal-to-noise ratio. In addition, the user must specify a maximum tolerable relative testing cost, say E_R^*, and a maximum tolerable Type I error, say E_U^*. Subject to $E_R \leq E_R^*$ and $E_U \leq E_U^*$, the search program determines, for various group sizes, the values of α_1 and α_2 that maximize power (E_A). From the program output, the group size (and corresponding α_1 and α_2 levels) which gives the greatest power may then be selected. The basic steps comprising the search algorithm are outlined in Table 1.

The logic of the search is based on the premise that power is an increasing function of testing cost. That is, the more runs that are invested, the greater is the power. Consequently, we can first determine α_1 independently of α_2 since E_R is a function of α_1 alone. Moreover, we find (if possible) the α_1 such that $E_R = E_R^*$. Once α_1 has been determined, it is not difficult to determine the optimal value of α_2 subject to $E_U \leq E_U^*$. We remark that a "solution" in which E_R is strictly less than E_R^* would be counterintuitive.

Step 1.	Input values for K, k, and Δ/σ .
Step 2.	Input maximum tolerable values for E_R and E_U.
Step 3.	Assume $\underline{\beta}(0)$ case and g=2.
Step 4.	Determine α_1 so that E_R attains maximum allowable value.
Step 5.	For the α_1 determined in Step 4, determine the α_2 that maximizes E_U subject to constraint specified in Step 2.
Step 6.	Calculate E_A and E_U for given $GS(g,\alpha_1,\alpha_2)$ strategy.
Step 7.	Repeat Steps 4, 5, and 6 as long as $g \leq \min(8, K/2)$.
Step 8.	Reset g=2 and repeat Steps 4 through 7 for $\underline{\beta}(\lceil k/2 \rceil)$ case.

Table 1. Outline of Two-Stage Group Screening Search Algorithm

4. USER'S GUIDE

In this section we discuss several key aspects of the search program, detail the input considerations and requirements, describe the various program modules, and present some sample printout. In this section, numbers appearing in brackets refer to line numbers in the program listing which is provided in Section 5.

4.1 IMSL Subprograms

The search program makes direct use of three IMSL (The International Mathematical and Statistical Library) subprograms. These are MDTN, MDSTI, and MDBIN. The IMSL subroutine MDTN calculates a non-central t probability and MDBIN calculates a binomial probability. The IMSL routine MDSTI is used to obtain inverse values of Student's t probability distribution function.

Because of copyright regulations, however, the source listings for each of these routines (and any other IMSL routines called by these routines) are not included with the search program listing given in Section 5. Therefore, to apply the search program the user must either have calling access to the ISML package of subprograms or provide equivalent alternative routines. The Appendix contains additional documentation of the MDTN, MDSTI, and MDBIN routines.

4.2 ERRSET Subroutine

The ERRSET subroutine is part of the Extended Error Handling Facility provided by IBM. As called by the search program, see [22], ERRSET suppresses all underflow messages and allows for an unlimited

number of such occurrences. The standard corrective action for underflow, i.e., set result register to zero, is unchanged. Moreover, no traceback is printed. Underflow can arise in the search routine when tail probabilities that are essentially zero must be calculated.

4.3 Input Considerations

Application of the two-stage group screening search routine requires a minimal amount of effort on the part of the user. Only two input data (control) cards must be prepared. Furthermore, to simplify the input process format-free input is assumed.

The first input data card must contain values for K, k, and Δ/σ, in that order. The second input data card must contain values for E_R^* and E_U^*, in that order. Furthermore, the user must express E_R^* and E_U^* as percentages.

The values of K, k, and Δ/σ are limited only by their output formats, see [43/44]. In addition, for ease of calculation[1], k is assumed to be even. If k is odd, the program automatically redefines k as k+1 and prints a message to that effect.

The search program also contains several checks for invalid input parameters, see [27/31]. The following error condition codes are used:

Condition Code	Description
1	$E_R^* \geq 100$. In this case, we recommend the use of a PB design (i.e., g=1) for screening.
2	$E_R^* \leq 0$.

[1] If k is even, there are an equal number of positive (+Δ) and negative (-Δ) active effects in the $\underline{\beta}([k/2])$ case. In this case the search program can exploit the existing symmetry in order to substantially reduce the extent of the required performance calculations. See Technical Report No. 113-9 for further details.

Condition Code	Description
3	$E_U^* \leq 0$ or $E_U^* \geq 100$.
4	$k \geq K$.
5	$\Delta/\sigma \leq 0$.

4.4 Program Modules

A brief description of each program module in the search program is provided in Table 2. This description provides information about the purpose of the module, its location in the source listing, and other modules which call and are called by it. The PER subprogram is the key subroutine that evaluates performance. PER can evaluate the performance of any GS(g, α_1, α_2) strategy.

4.5 An Example

Table 3 contains sample computer printout from the search routine for the case K = 60, k = 8, Δ/σ = 2, E_R^* = 50%, and E_U^* = 5%. Asterisks appearing in the printout signify a group size in which the value of E_R at α_1 = 0 equals or exceeds E_R^*. In Table 3, for instance, when g = 2 the first-stage PB experiment requires 32 runs, thus leaving no runs available for the second-stage follow-up experiment. Asterisks, therefore indicate that E_R^* is out of range for the given group size.

The search routine is based on the performance results derived in Technical Report No. 113-9. These results, however, are only applicable when K is a multiple of g. Therefore, for group sizes where this restriction is not met, the program redefines K as the nearest multiple of g and denotes the new value as KSTAR. Performance results are then calculated as if there are KSTAR factors to be screened.

PROGRAM MODULE	LOCATION	PURPOSE	CALLS	CALLED BY
Main Program	[20/131]	Serves as the executive program for conducting the search	ERRSET PER	---
Subroutine PER	[133/164]	Determines performance characteristics (E_R, E_A, and E_U) of a GS(g, α_1, α_2) strategy	MDSTI R1 RM D1	Main Program
Subfunction C	[166/177]	Calculates C_R^N	---	R1 RM
Subfunction R1	[179/202]	Calculates the probability that an active factor reaches second stage	C MDTN	PER
Subfunction RM	[204/228]	Calculates the probability that an inactive factor reaches second stage	C MDTN	PER
Subfunction D1	[230/245]	Calculates the probability that an active factor is found to be significant in second stage given it reaches the second stage	MDBIN MDTN	PER
Subroutine ERRSET	External	See Section 4.2	*	*
Subroutine MDTN MDSTI MDBIN	External	See Section 4.1	*	*

Table 2. Program Modules

SEARCH RESULTS FOR B(0) CASE:

STRATEGY	GROUPS	KSTAR	POWER	TYPE I ERROR	RELATIVE TESTING COST
GS(2,*******,*******)	***	****	******	******	****
GS(3,0.00031,1.00000)	20	60	26.21	4.641	50.0
GS(4,0.00015,0.49499)	15	60	40.74	5.000	50.0
GS(5,0.00110,0.30370)	12	60	49.39	5.000	50.0
GS(6,0.04372,0.20769)	10	60	49.59	5.000	50.0
GS(7,0.00714,0.19609)	9	63	55.98	5.000	50.0
GS(8,0.00016,0.22686)	7	56	49.18	5.000	50.0

SEARCH RESULTS FOR B(4) CASE:

STRATEGY	GROUPS	KSTAR	POWER	TYPE I ERROR	RELATIVE TESTING COST
GS(2,*******,*******)	***	****	******	******	****
GS(3,0.00043,0.95345)	20	60	22.30	5.000	50.0
GS(4,0.00027,0.44386)	15	60	33.18	5.000	50.0
GS(5,0.00198,0.27931)	12	60	40.01	5.000	50.0
GS(6,0.05444,0.19763)	10	60	41.60	5.000	50.0
GS(7,0.01312,0.18467)	9	63	45.06	5.000	50.0
GS(8,0.00075,0.20804)	7	56	37.21	5.000	50.0

Table 3. Sample Computer Printout From Two-Stage Group Screening Search Routine When K=60, k=8, and Δ/σ=2. Maximum E_U Specified As 5% And Maximum E_R Specified As 50% (Equivalently, E(R)=32 Runs). Note: Power, Type I Error, and Relative Testing Cost Are Expressed As Percentages.

For example, in Table 3 for g = 7, the number of groups is taken as G = 9, and thus KSTAR = gG = 63. It seems reasonable that the corresponding results for g = 7 should be comparable to the true performance had the original K = 60 factors been partitioned as "evenly" as possible - i.e., 6 groups of size 7 and 3 groups of size 6.

An examination of Table 3 shows that the GS(7, .00714, .19609) strategy has greatest power in the $\underline{\beta}(0)$ case and the GS(7, .01312, .18467) strategy has greatest power in the $\underline{\beta}(4)$ case. Greater power, of course, is attained in the $\underline{\beta}(0)$ case, when no cancellation can occur.

The search routine presented and described in this report provides guidance in using and selecting a satisfactory GS strategy. The routine supplies the user with quantitative information needed to determine whether two-stage group screening is suitable for a particular application.

5. PROGRAM LISTING

```
      C  TWO-STAGE GROUP SCREENING SEARCH ROUTINE
      C
      C  ------------------------------------------------------------
      C  AUGUST 1983.  DESMATICS, INC.
      C                P.O. BOX 618
      C                STATE COLLEGE, PA  16801
      C  ------------------------------------------------------------
      C  USER MUST SUPPLY FOLLOWING TWO CONTROL CARDS (FREE FORMAT):
      C                CARD #1
      C  K      -   TOTAL NUMBER OF FACTORS TO BE SCREENED
      C  NACT   -   NUMBER OF ACTIVE(I.E.,NONZERO) FACTORS
      C  SNR    -   SIGNAL-TO-NOISE RATIO FOR ACTIVE FACTORS
      C                CARD #2
      C  RTC    -   MAXIMUM RELATIVE TESTING COST (PER CENT)
      C  T1E    -   MAXIMUM TYPE I ERROR (PER CENT)
      C  ------------------------------------------------------------
      C  PROGRAM IS BASED ON DESMATICS, INC. TECHNICAL REPORT NO.
      C  113-9.  ALSO SEE USER'S GUIDE, TECHNICAL REPORT NO. 113-10.
      C  ------------------------------------------------------------
            DIMENSION PR(7,5),IPR(7,2)
            INTEGER G,GS,GLIM
            CALL ERRSET (208,256,-1,1,1,0)
      C                                READ VALUES FOR K,NACT,SNR
            READ (5,*) K,NACT,SNR
      C                                READ VALUES FOR RTC,T1E
            READ (5,*) RTC,T1E
            IF (RTC .GE. 100.) STOP 1
            IF (RTC .LE. 0.) STOP 2
            IF (T1E .LE. 0. .OR. T1E .GE. 100.) STOP 3
            IF (NACT .GE. K) STOP 4
            IF (SNR .LE. 0.) STOP 5
            IPBK=K+5-MOD(K+1,4)
            PBK=IPBK
            ER=PBK*RTC/100
            IF1=0
            IF (MOD(NACT,2) .EQ. 0) GO TO 5
            IF1=1
            NACT=NACT+1
          5 CONTINUE
            WRITE (6,10)
         10 FORMAT ('1','TWO-STAGE GROUP SCREENING SEARCH ROUTINE')
            WRITE (6,15) K,NACT,SNR
         15 FORMAT ('0','FOR THE CASE: FACTORS =',I4,/,15X,'ACTIVE = ',
           *         I3,/,15X,'SIGNAL-TO-NOISE RATIO =',F4.1)
            IF (IF1 .EQ. 0) WRITE (6,20)
         20 FORMAT ('-')
            IF (IF1 .EQ. 1) WRITE (6,25)
         25 FORMAT ('0','NOTE: THE NUMBER OF ACTIVE FACTORS HAS BEEN',/,
           *         7X,'INCREMENTED BY ONE SO THAT IT IS EVEN')
            WRITE (6,30) RTC,ER,T1E
```

```
      30 FORMAT ('-',/,1X,'RELATIVE TESTING COST = ',F4.1,3X,
     *         '(EXPECTED NUMBER OF RUNS =',F6.1,')',/,
     *        1X,'MAXIMUM TYPE I ERROR = ',F5.2,//,1X,'NOTE: ',
     *         'POWER, TYPE I ERROR, AND RELATIVE TESTING COST ARE',
     *         ' EXPRESSED AS PERCENTAGES')
         GLIM=MINO(K/2,8)
         I=NACT
         DO 90 J=1,2
            IF (J .EQ. 2) I=NACT/2
C                        START LOOP OVER GROUP SIZES
            DO 60 GS=2,GLIM
               M=GS-1
               Z=FLOAT(K)/GS
               G=Z
               IF (Z-G .GT. .5) G=G+1
               KS=G*GS
               PBG=G+5-MOD(G+1,4)
               PBKS=KS+5-MOD(KS+1,4)
               GAM=100*PBG/PBKS
               IF (RTC .GT. GAM+350./PBKS) GO TO 35
               PR(M,1)=10
               PR(M,2)=10
               IPR(M,1)=1000
               IPR(M,2)=10000
               PR(M,3)=1000
               PR(M,4)=100
               PR(M,5)=100
               GO TO 60
      35       CONTINUE
C                        DETERMINE FIRST STAGE SIGNIFICANCE LEVEL
               A=0
               B=1
               FA=GAM
               FB=100+GAM
               SL1=(RTC-GAM)/100
               CALL PER (1,REA,REU,RER,KS,NACT,SNR,I,GS,SL1,SL2)
               F=RER
      40       IF (ABS(F-RTC) .LT. .025) GO TO 55
               IF (F .LT. RTC) GO TO 45
               B=SL1
               FB=F
               GO TO 50
      45       A=SL1
               FA=F
      50       CONTINUE
               SLOPE=(FB-FA)/(B-A)
               SL1=A+(RTC-FA)/SLOPE
               CALL PER (1,REA,REU,RER,KS,NACT,SNR,I,GS,SL1,SL2)
               F=RER
               GO TO 40
```

```
101.          55      PR(M,1)=SL1
102.                  IPR(M,1)=G
103.                  IPR(M,2)=KS
104.                  CALL PER (0,REA,REU,RER,KS,NACT,SNR,I,GS,SL1,.1)
105.      C                   DETERMINE SECOND-STAGE SIGNIFICANCE LEVEL
106.                  SL2=AMIN1(1.,T1E/(10*REU))
107.                  PR(M,2)=SL2
108.      C                   CALCULATE POWER FOR SPECIFIED STRATEGY
109.                  CALL PER (0,REA,REU,RER,KS,NACT,SNR,I,GS,SL1,SL2)
110.                  PR(M,3)=REA
111.                  PR(M,4)=REU
112.                  PR(M,5)=RER
113.          60      CONTINUE
114.                I=NACT-I
115.                  WRITE (6,65) I
116.          65      FORMAT ('-',//////,1X,'SEARCH RESULTS FOR B(',I2,') CASE:')
117.                  WRITE (6,70)
118.          70      FORMAT ('0',6X,'STRATEGY',11X,'GROUPS',4X,'KSTAR',4X,
119.            *           'POWER',4X,'TYPE I ERROR',4X,'RELATIVE TESTING COST')
120.                  WRITE (6,75)
121.          75      FORMAT ('+',6X,'________',11X,'______',4X,'_____',4X,
122.            *           '_____',4X,'___ - _____',4X,'_____ _____ ____')
123.                  DO 80 GS=2,GLIM
124.                     M=GS-1
125.          80         WRITE (6,85) GS,(PR(M,I),I=1,2),(IPR(M,I),I=1,2),
126.            *                      (PR(M,I),I=3,5)
127.          85      FORMAT (' ','GS(',I1,',',F7.5,',',F7.5,')',5X,I3,6X,I4,
128.            *           4X,F6.2,7X,F6.3,16X,F4.1)
129.          90      CONTINUE
130.                STOP
131.                END
132.      C
133.                SUBROUTINE PER(ICODE,REA,REU,RER,K,NACT,R,II,GS,SL1,SL2)
134.      C
135.      C           IF ICODE EQUALS 1 THEN ONLY RER RETURNED
136.      C
137.                INTEGER G,GS
138.                REAL CV(4)
139.                REAL*8 C,R1,RM,ES,RA,RB
140.                G=K/GS
141.                KNACT=K-NACT
142.                IBK=K+5-MOD(K+1,4)
143.                IPB=G+5-MOD(G+1,4)
144.                IDF1=IPB-G-1
145.                PB1=IPB
146.                RIDF1=IDF1
147.                CALL MDSTI(SL1,RIDF1,TVAL,IER)
148.                RA=R1(K,GS,NACT,II,R,TVAL,IDF1,PB1)
149.                RB=RM(K,GS,NACT,II,R,TVAL,IDF1,PB1)
150.                ES=NACT*RA+KNACT*RB
```

```
151.            RER=100.*(PB1+3.5+ES)/PBK
152.            IF(ICODE.EQ.1) RETURN
153.            DA=1.
154.            IF(SL2.EQ.1) GO TO 4
155.            DO 3 I=1,4
156.            RI=I
157.          3 CALL MDSTI(SL2,RI,CV(I),IER)
158.            PG=ES/K
159.            DA=D1(GS,G,PG,R,CV)
160.          4 CONTINUE
161.            REA=100.*DA*RA
162.            REU=100.*SL2*RB
163.            RETURN
164.            END
165.      C
166.            DOUBLE PRECISION FUNCTION C(N,R)
167.            INTEGER R
168.            C=0.0D0
169.            IF(R.GT.N) RETURN
170.            C=1.0D0
171.            IF(R.EQ.0.OR.R.EQ.N) RETURN
172.            J=R
173.            IF(R.GT.N/2) J=N-R
174.            DO 10 I=1,J
175.         10 C=C*(N-I+1)/(J-I+1)
176.            RETURN
177.            END
178.      C
179.            DOUBLE PRECISION FUNCTION R1(K,GS,NACT,I,R,TVAL,IDF1,PB1)
180.            INTEGER GS,GSJ,GSJL
181.            REAL*8 C,PJL
182.            KNACT=K-NACT
183.            I1=I-1
184.            NACTI=NACT-I
185.            D=SQRT(PB1)*R
186.            L1=MIN0(GS,I)
187.            R1=0.D0
188.            DO 10 J=1,L1
189.            J1=J-1
190.            GSJ=GS-J
191.            L2=MIN0(GSJ,NACTI)+1
192.            DO 20 L=1,L2
193.            GSJL=GSJ-L+1
194.            PJL=C(I1,J1)*C(NACTI,L-1)*C(KNACT,GSJL)
195.            DP=D*(L-J-1)
196.            CALL MDTN(TVAL,IDF1,DP,A1,IER)
197.            CALL MDTN(TVAL,IDF1,-DP,A2,IER)
198.         20 R1=R1+PJL*(2.D0-A1-A2)
199.         10 CONTINUE
200.            R1=R1/C(K-1,GS-1)
```

```
      RETURN
      END
C
      DOUBLE PRECISION FUNCTION RM(K,GS,NACT,I,R,TVAL,IDF1,PB1)
      INTEGER GS,GSJ,GSJL
      REAL*8 C,PJL
      KNACT=K-NACT
      I1=I-1
      NACTI=NACT-I
      KNACT1=KNACT-1
      D=SQRT(PB1)*R
      L1=MIN0(GS,I+1)
      RM=0.D0
      DO 10 J=1,L1
      J1=J-1
      GSJ=GS-J
      L2=MIN0(GSJ,NACTI)+1
      DO 20 L=1,L2
      GSJL=GSJ-L+1
      PJL=C(I,J1)*C(NACTI,L-1)*C(KNACT1,GSJL)
      DP=D*(L-J)
      CALL MDTN(TVAL,IDF1,DP,A1,IER)
      CALL MDTN(TVAL,IDF1,-DP,A2,IER)
   20 RM=RM+PJL*(2.D0-A1-A2)
   10 CONTINUE
      RM=RM/C(K-1,GS-1)
      RETURN
      END
C
      REAL FUNCTION D1(GS,G,PG,R,CV)
      INTEGER GS,G,G1,S1
      REAL CV(4)
      G1=G-1
      D1=0.
      DO 10 J=1,G
      CALL MDBIN(J-1,G1,PG,PS,PJJ,IER)
      S1=GS*J+1
      N2=S1+4-MOD(S1,4)
      IDF2=N2-S1
      DP=SQRT(FLOAT(N2))*R
      CALL MDTN(CV(IDF2),IDF2,DP,A1,IER)
      CALL MDTN(CV(IDF2),IDF2,-DP,A2,IER)
   10 D1=D1+PJJ*(2.-A1-A2)
      RETURN
      END
```

6. APPENDIX

The following four documentation pages were taken from the IMSL Library Reference Manual, IMSL LIB-0008, June 1980.

IMSL ROUTINE NAME - MDBIN

PURPOSE - BINOMIAL PROBABILITY DISTRIBUTION FUNCTION

USAGE - CALL MDBIN (K,N,P,PS,PK,IER)

ARGUMENTS
- K - NUMBER OF SUCCESSES (INPUT)
- N - NUMBER OF BERNOULLI TRIALS (INPUT)
- P - PROBABILITY OF SUCCESS ON EACH TRIAL (INPUT)
- PS - PROBABILITY THAT THE NUMBER OF SUCCESSES IS K OR LESS (OUTPUT)
- PK - PROBABILITY THAT THE NUMBER OF SUCCESSES IS EXACTLY K (OUTPUT)
- IER - ERROR PARAMETER. (OUTPUT)
  TERMINAL ERROR
  IER = 129 INDICATES THAT K WAS SPECIFIED LESS THAN ZERO OR GREATER THAN N.
  IER = 130 INDICATES THAT P WAS SPECIFIED GREATER THAN 1 OR LESS THAN 0.

PRECISION/HARDWARE - SINGLE/ALL

REQD. IMSL ROUTINES - UERTST,UGETIO

NOTATION - INFORMATION ON SPECIAL NOTATION AND CONVENTIONS IS AVAILABLE IN THE MANUAL INTRODUCTION OR THROUGH IMSL ROUTINE UHELP

Algorithm

MDBIN computes the probability, PK, of obtaining exactly K successes and PS, the probability of obtaining K or less successes in N Bernoulli trials, where P is the probability of success (Q=1-P is probability of failure) on each trial.

The following recursive relationship is used

$$Pr(X=J)=Pr(X=J-1)\cdot(N+1-J)\cdot P/(J\cdot Q)$$

Then

$$Pr(X \text{ not greater than } K) = \sum_{J=0}^{K} Pr(X=J)$$

To eliminate the possibility of underflow, PS and PK are computed forward from 0 if K is not greater than N*P and backward from N otherwise. The scaled quantities PS and PK, are built recursively using an initial value of EPS, where EPS is the smallest positive machine number. These quantities are rescaled by a multiplicative factor of Q^N*EPS if forward computation is done and P^N*EPS if backward computation is performed.

For the special case P=0.,

$$PK = \begin{cases} 0 & \text{if } K \neq 0 \\ 1 & \text{if } K = 0 \end{cases}$$

PS = 1

IMSL ROUTINE NAME - MDSTI

PURPOSE - INVERSE OF A MODIFICATION OF STUDENTS T PROBABILITY DISTRIBUTION FUNCTION

USAGE - CALL MDSTI (Q,F,X,IER)

ARGUMENTS

Q - INPUT PROBABILITY IN THE EXCLUSIVE RANGE (0,1). (THE SUM OF THE AREAS (EQUAL) IN BOTH TAILS OF THE T DISTRIBUTION.)

F - INPUT DEGREES OF FREEDOM FOR T DISTRIBUTION. F MUST BE GREATER THAN ZERO.

X - OUTPUT VALUE SUCH THAT THE PROBABILITY OF THE ABSOLUTE VALUE OF T BEING GREATER THAN X IS Q.

IER - ERROR PARAMETER. (OUTPUT)
WARNING ERROR
IER = 35 INDICATES OVERFLOW WOULD HAVE OCCURRED. X IS SET TO MACHINE INFINITY.
TERMINAL ERROR
IER = 129 INDICATES THAT F (DEGREES OF FREEDOM) IS LESS THAN OR EQUAL TO ZERO.
IER = 130 INDICATES THAT Q IS OUT OF THE EXCLUSIVE RANGE (0,1).
IER = 131 INDICATES THAT AN ERROR OCCURRED IN IMSL ROUTINE MDNRIS, THE INVERSE NORMAL PROBABILITY DISTRIBUTION FUNCTION.
IER = 132 INDICATES THAT CONVERGENCE WAS NOT ACHIEVED. THIS CAN ONLY OCCUR FOR F LESS THAN 2.0 .

PRECISION/HARDWARE - SINGLE/ALL

REQD. IMSL ROUTINES - MDNRIS,MERFI,UERTST,UGETIO

NOTATION - INFORMATION ON SPECIAL NOTATION AND CONVENTIONS IS AVAILABLE IN THE MANUAL INTRODUCTION OR THROUGH IMSL ROUTINE UHELP

REMARKS NOTE THAT MDSTI DOES NOT PROVIDE THE ACTUAL T INVERSE. FOR P EQUAL TO THE PROBABILITY THAT A STUDENTS T RANDOM VARIABLE IS LESS THAN X, THAT INVERSE CAN BE OBTAINED BY THE FOLLOWING RULES.

A. FOR P IN THE RANGE (0.0,0.5), CALL MDSTI WITH Q = 2*P AND NEGATE THE RESULT X.
B. FOR P IN THE RANGE (0.5,1.0), CALL MDSTI WITH Q = 2*(1-P).

Algorithm

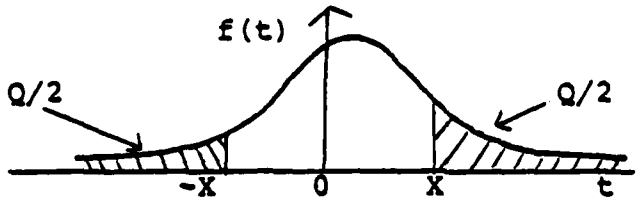
This subroutine computes X such that

$$Q = 1 - \int_{-\infty}^{X} f(t)\,dt$$

where

$$f(t) = \frac{\Gamma\left(\frac{F+1}{2}\right)}{\sqrt{\pi F}\,\Gamma\left(\frac{F}{2}\right)} \cdot \left(1 + \frac{t^2}{F}\right)^{-\frac{F+1}{2}} \qquad \text{for } -\infty < t < \infty$$

See the following figure.



See reference:

1. Hill, G.W., "Algorithm 396, Student's t-quantiles", CACM, 13(10)1970, 619-620.

2. Roe, G.M., Programs for the Incomplete Beta and Gamma Functions and their Inverses, General Electric Technical Information Series, January, 1969.

Accuracy

Test results indicate 5 significant digit accuracy. In addition, Table 26.10, p. 999 of the Handbook of Mathematical Functions, edited by M. Abramowitz and I.A. Stegun, was duplicated by MDSTI.

Example

In this example X is calculated such that the probability of the absolute value of a random variable following Student's t distribution with two degrees of freedom being greater than X is 0.8.

Input:

```
INTEGER  IER
REAL     Q,F,X
Q    = 0.8
F    = 2.

CALL MDSTI (Q,F,X,IER)
```

Output:

```
X    = 0.28867
IER = 0
```

IMSL ROUTINE NAME - MDTN

PURPOSE - NON-CENTRAL T PROBABILITY DISTRIBUTION FUNCTION

USAGE - CALL MDTN (TVAL,IDF,D,P,IER)

ARGUMENTS
- TVAL - INPUT VALUE TO WHICH INTEGRATION IS PERFORMED.
- IDF - INPUT. INTEGER DEGREES OF FREEDOM.
- D - INPUT. NON-CENTRALITY PARAMETER.
- P - OUTPUT. PROBABILITY THAT A RANDOM VARIABLE DISTRIBUTED AS T WITH NON-CENTRALITY PARAMETER D IS LESS THAN OR EQUAL TO TVAL.
- IER - ERROR PARAMETER. (OUTPUT) TERMINAL ERROR
  IER = 129, INDICATES DEGREES OF FREEDOM OUT OF RANGE. IDF MUST BE GREATER THAN ZERO.

PRECISION/HARDWARE - SINGLE/ALL

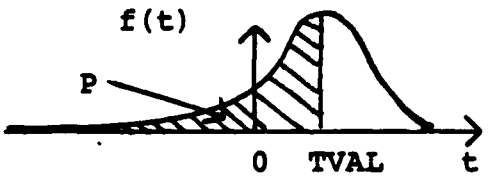
REQD. IMSL ROUTINES - MDNOR,MDTNF,MERRC=ERFC,UERTST,UGETIO

NOTATION - INFORMATION ON SPECIAL NOTATION AND CONVENTIONS IS AVAILABLE IN THE MANUAL INTRODUCTION OR THROUGH IMSL ROUTINE UHELP

Algorithm

MDTN calculates the probability P that a random variable Z which follows the non-central t distribution, with non-centrality parameter D, is less than or equal to TVAL.

Letting f(t) be the non-central T probability density function with non-centrality parameter D, the indicated P value is returned.



See reference:

Owen, D.B., "A special case of the bivariate non-central t-distribution", Biometrika, 52, 1965, 437-446.

Example

This example illustrates use of MDTN with a positive non-centrality parameter, D=6.0.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 113-10	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Two-Stage Group Screening Search Routine: Program Description and User's Guide		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Carl A. Mauro Kevin C. Burns		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0650
9. PERFORMING ORGANIZATION NAME AND ADDRESS Desmatics, Inc. P.O. Box 618 State College, PA 16801		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-467
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		12. REPORT DATE September 1983 13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this report is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Factor Screening
Group Screening
Search Program

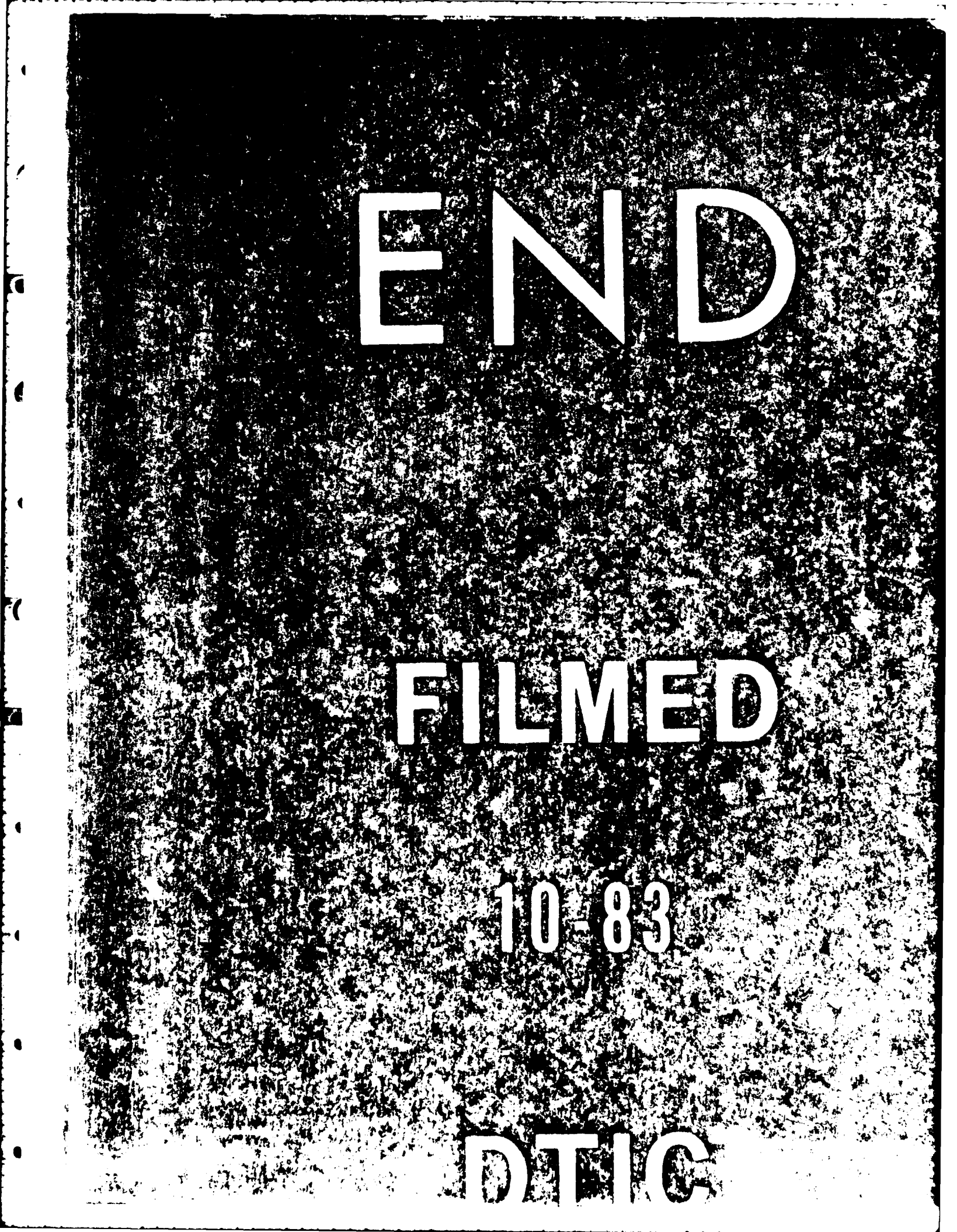
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A computer-aided search program has been developed to help select a satisfactory two-stage group screening strategy. The program is easy to use and provides practical insight into the use and selection of a two-stage group screening strategy. This report describes the program and serves as a user's guide. A listing of the program is included.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)


END
FILMED
10-83
DTIC